



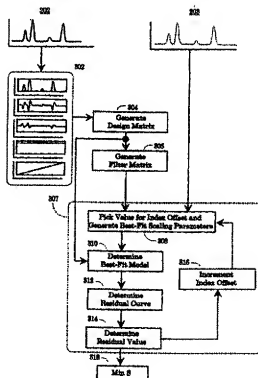
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup>:</b> <b>G01N 30/86, G06F 15/20</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/39347</b>  <b>(43) International Publication Date:</b> 23 October 1997 (23.10.97)
<b>(21) International Application Number:</b> PCT/US97/06135 <b>(22) International Filing Date:</b> 14 April 1997 (14.04.97) <b>(30) Priority Data:</b> 08/631,446 12 April 1996 (12.04.96) US <b>(71) Applicant:</b> WATERS INVESTMENTS LIMITED [US/US]; 34 Maple Street, Milford, MA 01757 (US). <b>(72) Inventor:</b> GORENSTEIN, Marc, V.; 313 Brooklin Street, Needham, MA 02192 (US). <b>(74) Agent:</b> JANIUK, Anthony, J.; Waters Investments Limited, Legal Dept., 34 Maple Street, Milford, MA 01757 (US).	<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>	

**(54) Title:** CHROMATOGRAPHIC PATTERN ANALYSIS SYSTEM EMPLOYING CHROMATOGRAPHIC VARIABILITY CHARACTERIZATION

**(57) Abstract**

A Chromatographic Pattern Analysis System (CPAS) (100) determines chromatographic variability due to a plurality of sources without requiring identification or characterization of peaks or other chromatographic features, receives data indicative of a standard chromatogram and a first sample chromatogram generated from a first mixture by a High Pressure Liquid Chromatography (HPLC) device (102) and data indicative of a plurality of additional sample chromatograms generated by the HPLC device (102) from a plurality of different mixtures. The CPAS (100) generates from the standard chromatogram a plurality of sets of chromatographic variability data, each set being indicative of a different effect of the chromatographic variability of the HPLC. The standard chromatogram is modified as a function of the variability data, and a residual value, indicative of a difference between the modified standard chromatograms and the first sample chromatogram is generated. Residual values are generated for the additional sample chromatograms and are used to determine differences between the corresponding mixtures and the first mixture.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Lucembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroun	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## CHROMATOGRAPHIC PATTERN ANALYSIS SYSTEM EMPLOYING CHROMATOGRAPHIC VARIABILITY CHARACTERIZATION

5

### AUTHORIZATION

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears  
10 in the Patent and Trademark Office patent file or records,

### FIELD OF THE INVENTION

This invention relates in general to the field of analysis of chromatograms and more particularly, though not limited to, determining chromatographic variability in  
15 order to increase the accuracy of chromatographic analysis.

### BACKGROUND OF THE INVENTION

Chromatography is a technique widely used in the analysis of multi-component substances. In chromatography, a liquid or gas, of known or unknown composition, is  
20 injected into a chromatograph which generates a chromatogram in the form of a two dimensional graph in which absorptivity of the injected liquid, or conductivity of the injected gas or some other physical response is plotted against time. The absorptivity of the liquid or conductivity of the gas with respect to time as it passes through the chromatograph is indicative of the composition of the liquid or gaseous mixture.

Common uses of chromatography include quality control, in which manufactured substances are analyzed to verify the composition, and qualitative and quantitative analysis in which chromatograms derived from unknown substances are generated to analyze and determine the composition of the substance. In a quality control application, a chromatogram of a known and desired substance is generated and compared to the chromatogram of the manufactured substance. In qualitative and quantitative analysis applications, one or more chromatograms are generated of the unknown substance in an attempt to identify the components or quantify the amounts of each of the components of the substance.

In either of the above applications, the chromatogram(s) must be analyzed to determine similarities or differences with other chromatograms. Typically such analysis requires analysis and comparison of the peaks, including the retention time, height and area of peaks between one chromatogram with those of another. In order to perform such a comparison, a method of identifying the peaks to be compared must be developed and optimized, and then a method to compare the peaks and other above identified aspects of the chromatograms must be developed. Principal Component Analysis (PCA) is one such technique and is referred to by G. Malmquist and R. Danielsson in a paper entitled "Alignment of Chromatographic Profiles for Principal Component Analysis: A Prerequisite for Fingerprinting Methods", *Journal of Chromatography A*, 687 (1994) 71-88. As described by Malmquist and Danielsson, retention times of selected peaks are used to align corresponding chromatograms. Once they are aligned, the absorbances are themselves compared. Another technique is described by J.P. Mason et al. in an article entitled "A Novel Algorithm for

Chromatogram Matching in Qualitative Analysis", *Journal of High Resolution Chromatography*, v. 15, pp. 539-547 (August 1992) which describes an automated chromatographic matching technique which compares only peak heights, areas and retention times.

- 5           Chromatographic analysis must also take into account variability introduced by the chromatograph such as baseline drift, retention time wander and concentration change. Such variations of the chromatograph are manifested as variations in the chromatograms generated by the chromatograph and further complicate the analysis by requiring the analysis to take into account those several variations. In the above
- 10   referenced paper by Malmquist and Danielsson, a technique is described for compensating for chromatographic variability in the context of enhancing chromatographic analysis by PCA.

- Known techniques for chromatographic analysis such as described by Malmquist and Danielsson and Mason et al. typically require the steps described above
- 15   of specifying a method of identifying peaks to be compared, specifying a method of comparing the various aspects of the peaks, and then actually performing the comparisons, while taking into account the effects of chromatographic variability. Although computerized techniques such as those described by Mason are helpful in performing such tasks, many known techniques continue to be time consuming,
- 20   sometimes tedious and require the skills of highly trained personnel.

It is accordingly an object of the present invention to provide a system for chromatographic analysis which compensates for chromatographic variability and

which performs chromatographic analysis without requiring the characterization of peaks or other chromatographic features as required by known techniques.

#### SUMMARY OF THE INVENTION

5 In accordance with the primary object of the invention, a chromatographic analysis system employs a chromatogram alignment procedure which characterizes variability of a chromatograph. In accordance with a further object of the invention, the chromatographic analysis system employs the characterized chromatographic variability in the comparison of chromatograms. Advantageously, the chromatographic analysis system of the present invention of the present invention compares  
10 chromatograms as patterns and does not require the characterization of peaks or the identification of the accompanying peak lift-off or touch down points. Moreover, peak heights, areas or retention times need not be computed. The phrase "characterization of peaks" is intended to refer to the process of peak integration which requires identification of the baseline by identification of the lift and touchdown points of the  
15 peak in question and then actual integration of the peak. The chromatographic pattern analysis system of the present invention applies to chromatograms obtained by isocratic or gradient chromatographic separations. In isocratic chromatography, the composition of the solvent is held constant throughout the chromatographic  
20 separation. In gradient chromatography, the composition of the solvent is varied in a predetermined way to obtain enhanced control over the retention times of compounds.

In a first aspect, a chromatographic analysis system operating in accordance with the principles of the present invention determines differences between a standard

- chromatogram and a sample chromatogram. The standard chromatogram is represented by a set of standard data points, indicative of the standard chromatogram over a selected elution time range and the sample chromatogram is represented by a set of sample data points indicative of the sample chromatogram over the same elution
- 5 time range or an elution time range offset by a fixed amount. The standard data points and the sample data points are each generated by sampling each respective chromatogram at a fixed rate. The chromatographic analysis system then generates a plurality of sets of chromatographic variability data points from the standard data points, each of the sets of chromatographic variability data points being indicative of
- 10 effects of a predetermined source of chromatographic variability on the standard chromatogram. The system also generates a set of modified standard data points, which correspond to the standard data points modified as a function of the chromatographic variability data points, to model chromatographic variability of a chromatograph which generates said chromatograms.
- 15 The system described above may be used with standard and sample chromatograms generated from the same mixtures in order to determine variability of the chromatograph. In addition, the system may be used to determine similarities or differences between the standard chromatogram and sample chromatograms from different or unknown mixtures. To assist in such a comparison, the system, in a
- 20 second aspect, generates residue values which are indicative of differences between the standard chromatogram and the sample chromatograms. The residue value generated from a comparison of the standard to sample chromatograms of the same mixture may then be compared to residue values generated from a comparison of standard and

sample chromatograms of different mixtures to determine whether the mixtures corresponding to the sample and the standard chromatograms are the same or different mixtures.

The chromatographic variability data points derived from the standard chromatogram allow the standard chromatogram to be modified to reflect a broad range of possible chromatograms that differ from the standard only by the effects of chromatographic variability.

The system uses the chromatographic variability data points derived from the standard to find a model chromatogram that most closely matches the sample chromatogram. In this sense, the system measures chromatographic variability reflected in the sample by modifying the standard chromatogram.

Scale factors generated as a function of the chromatographic variability data points, that describe this "best-fit" model chromatogram are one measure of chromatographic variability between the standard and sample. The resultant difference between the model and the sample is another measure. These are both advantageously measures of chromatographic variability in the sense that the system determines how much variability has to be applied to the standard chromatogram (through the use of the chromatographic variability data points) to get the model based on the standard chromatogram to match the sample chromatogram.

The chromatographic variability data points derived from the standard chromatogram need only be generated once for each standard and can advantageously be used repeatedly in determining the variabilities between the standard and multiple sample chromatograms.



Embodiments operating in accordance with the principles summarized above advantageously provide an accurate determination of chromatographic variability without requiring identification or characterization of particular peaks or other features of the chromatograms in question. When chromatograms are from the same or similar mixtures chromatographic variability due to baseline drift, retention time wander and concentration change are measured and removed regardless of the particular features of the chromatogram in question and without additional calibration of the chromatograph or addition of reference compounds to the mixtures in question. In a preferred embodiment, the chromatographic analysis system removes chromatographic variability due to concentration change, retention time offset, retention time stretch, baseline offset and baseline slope. In other embodiments the system may remove chromatographic variability due to only one of the above mentioned sources of chromatographic variability of different subsets of such sources.

These and other features and advantages of the present invention may be better understood by considering the following detailed description of certain preferred embodiments of the invention. In the course of this description, reference will be made to the attached drawings.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Fig. 1 is a block diagram of a preferred chromatographic analysis system coupled to an HPLC device;

Fig. 2 is a graph showing a standard chromatogram and a sample chromatogram which are used as inputs to illustrate operation of a preferred embodiment;

Fig. 3 is a flowchart showing the operation of a preferred embodiment;

5 Figs. 4(a) - 4(e) are graphs showing chromatographic variability data points generated by a preferred embodiment as a function of the standard chromatogram of Fig. 2;

Fig. 5 is a block diagram of an alternative use of the system shown in Fig. 1; and

10 Figs. 6(a) - 6(d), and 7(a) - 7(d), are graphs illustrating operation of a preferred embodiment for different offset values.

## DETAILED DESCRIPTION

Fig. 1 of the drawings shows a schematic diagram of a preferred  
15 Chromatographic Pattern Analysis System (CPAS) 100 coupled to receive data from a High-Pressure Liquid Chromatography (HPLC) system 102. Such a system is available from Waters Corporation, Milford, Massachusetts 01757 and preferably includes a pump under the trade name Waters™ 626 which continually pumps solvent through an injector, column, and detector, into a waste bottle. Such a system  
20 advantageously accommodates a wide variety of applications including high-resolution protein purification, peptide mapping, nucleic acid isolations, purification and analysis of oligosaccharides and analysis of mycolic acid from mycobacteria. The injector injects a sample mixture to be separated by the column into the solvent stream

- generated by the pump. Preferably the injector takes the form of a Waters™ 717 plus Autosampler which may be programmed via the chromatography manager described below to perform a variety of functions including automated routines for automatically adding, mixing and injecting samples. One example of a chromatographic column
- 5 which separates the components of the injected mixture is also available from Waters Corp. under the trade name Waters™ Delta-Pak™ C<sub>18</sub> steel column. Such a column has an inner diameter of 3.9 mm and a length of 150mm, packed with 5 micron size beads having a 300 Angstrom pore size. The detector measures absorptivity of an eluent stream exiting the column and digitizes the measured absorbance. Such a
- 10 detector is also available from Waters Corp. under the trade name Waters™ 486 Tunable UV Absorbance Detector and advantageously provides detection range from 190-600 nanometers (nm), a bandwidth of 8nm with an accuracy of  $\pm 2$ nm and a reproducibility of  $\pm 0.25$  nm. The CPAS 100 is preferably implemented on a data
- 15 station which controls the operation of the HPLC 102 and accepts the digitized output of the detector. Preferably such a data station takes the form of a PC based computer configured to execute the Windows 3.1 operating system available from Microsoft Corp., Redmond, Washington, and application programs compatible with Windows 3.1 including the Millennium 2010 Chromatography Manager which implements a chromatographic data management system to provide control of operation of the
- 20 HPLC device including the ability to program, document and link results derived from analyses performed by the device. Such a chromatography manager advantageously provides a relational database to facilitate the organization, storage and retrieval of

results generated by the HPLC 102 and a Graphical User Interface (GUI) to control data acquisition and other system operational functions.

Other types of chromatography devices may also be used in conjunction with CPAS 100 including other types of liquid chromatographs as well as gas

5 chromatographs. CPAS 100 requires data representative of chromatograms in the form of an analog voltage signal, where typically 1 volt equals 1 absorbance unit (AU) or in the form of a digitized signal. Typically signals are sampled and digitized at the rate of once per second. Chromatographic peaks are typically 30 to 60 seconds wide as measured from the lift-off to touch-down point of an isolated peak.

10 As seen diagrammatically in Fig. 1, CPAS 100 generates by comparison with a standard chromatogram, via Chromatographic Variability Measurement module 104, a residual value  $S^2$  for each sample chromatogram. As seen in Fig. 1, the Standard Chromatogram and Sample Chromatogram 1 are each generated from the same mixture. A plurality of additional chromatograms (seen as Sample Chromatograms 2,  
15 3, . . . N) may be generated from Mixtures 2, 3 . . . N and transferred to the Chromatographic Identification module 106 of CPAS 100 for the purpose of determining the similarities or differences between these chromatograms and the standard. The CPAS 100 may be employed for a variety of purposes including quality control in which sample chromatograms 2, 3, . . . , N which correspond to mixtures 2,  
20 3, . . . , N are compared to the standard chromatogram to determine differences between manufactured mixtures 2, 3, . . . , N and mixture 1 which represents a mixture having a desired composition. The CPAS 100 may also be employed for a number of other applications including mycolic acid analysis and tryptic mapping.

Fig. 2 of the drawings illustrates the standard chromatogram and sample chromatogram 1 of Fig. 1, which will be used to illustrate operation of a preferred embodiment. In Fig. 2, the standard chromatogram is shown in solid line 202 and sample chromatogram 1 is shown in dotted line 203. In the following explanation, various points on standard chromatogram 202 are referenced by even reference numbers 204-210 and points on sample chromatogram 203 are referenced by odd reference numbers 205-211. As seen in Fig. 2, standard and sample chromatograms 202 and 203 each include a plurality of peaks 204-211. As also seen in Fig. 2, the standard and sample chromatograms differ in a number of respects even though they are both obtained from the same mixture. For instance, the peaks of chromatogram 203 are shifted and stretched in time from the corresponding peaks of chromatogram 202. Moreover, the baseline of chromatogram 203 is shifted and sloped upward from chromatogram 202. In addition, the height of the peaks of chromatogram 203 differ from the height of corresponding peaks in chromatogram 202. As is known, such differences occur because the properties of the chromatograph itself differ slightly between injections of the same mixture. For instance, solvent flow rates in the chromatograph may change from day to day or from instrument to instrument, and baselines may drift. Table 1 below lists five common instrumental variations that affect chromatograms. In Table 1 below, the variation is listed in the leftmost column with the model value that measures the magnitude of the variation, the units of the model value and the physical sources of the variation listed in the respective columns to the right.

Variation	Model Value	Units	Examples of Physical Origin
Concentration ratio	$s$	none	Changes in sample or injector
Retention time offset	$t_0$	sec	Change in delay volume
Time scale expansion or contraction	$r$	sec/sec	Change in pump flow rate
Change in baseline offset	$b_0$	AU	Thermal drift in the detector or eluent
Change in baseline drift	$b_1$	AU/sec	Thermal drift in the detector or eluent

**Table 1**

- CPAS 100 estimates each of the model values listed above from a comparison of the standard chromatogram with sample chromatogram 1. First, the five model values are determined to account for the difference between the two chromatograms due to chromatographic variability (referred to as "model variability"). If there is no model variability, then the model values are  $s = 1$ ,  $r = 0$ ,  $t_0 = 0$ ,  $b_1 = 0$  and  $b_0 = 0$ .
- Once the model variability is determined, any remaining differences, termed residual variability, are determined. As used herein, the term "residual variability" is intended to refer to differences between two chromatograms which do not arise from chromatographic variability which is described by the model defined in Table 1. The model variability and residual variability are then employed to enhance the analysis of chromatograms from different mixtures.

Before data representative of chromatograms 202 or 203 is submitted to CPAS 100 for analysis, an elution time range, also referred to as a comparison range, of the

standard chromatogram, represented by the horizontal axis of the graph of Fig. 2 must be selected. The elution time range is chosen to include the peaks of interest of the standard and sample chromatographs to be compared. Preferably there is no restriction on the length of the elution time range, as long as the elution time range contains at least two peaks. Moreover, there is no restriction as to whether the peaks in the elution time range are of known or unknown compounds. For example, no peaks from a reference compound need to be included in the elution time range for the CPAS 100 to work. With two or more peaks, relative retention time shift and stretch, as well as baseline model values and concentration change can be determined. With only a single peak, retention time stretch cannot be determined. Accordingly, proper operation of the system requires inclusion of at least two peaks in the comparison range.

In the example of Fig. 2, the selected comparison range comprises four hundred points starting from point 21 and ending at point 420. In the explanation which follows, the comparison range is designated by an index  $i_N$  in which  $i_{start}$  designates the first point in the comparison range (point 21) and  $i_{stop}$  designates the last point in the comparison range (point 420). In this example, the comparison range includes  $N = 400$  points. Moreover, in the following explanation, vectors and matrices are designated by **bold characters**.

In addition to a comparison range, an offset range, expressed in units of the sampling index by the parameter  $K$ , must be selected. The offset range is indicative of a maximum possible retention time offset and specifies a range from  $-K$  to  $+K$  over which the comparison between the standard and sample chromatograms is varied

Thus, the standard chromatogram is compared to  $2K + 1$  overlapping,  $N$ -point regions of the sample chromatogram. A particular offset within the offset range is specified by a value  $k$  termed an offset index. In practice, the value  $2K + 1$  is often comparable to the width of a chromatographic peak. In the example which follows,  $K = 10$ , so offsets will range from  $k = +10$  to  $k = -10$  indices.

Fig. 3 of the drawings is a flow diagram showing the steps performed by a preferred embodiment to determine chromatographic variability from standard and sample chromatograms 202 and 203. In Fig. 3, data from standard chromatogram 202 is used to generate five chromatographic variability curves which are each embodied by a corresponding set of data, herein termed "chromatographic variability data points" shown in graphical form at 302 and in greater detail in Figs 4(a) - 4(e). Associated with each curve is the corresponding model value listed in Table 1 above. These chromatographic variability curves embody the notion of a set of chromatographic variability data points.

The data points seen in Fig. 4(a) are generated by subtracting from the standard chromatogram, the minimum absorbance measured within the selected elution time range. The minimum absorbance measured within the selected elution time range can be found by collecting and sorting all the absorbance values from within that time range, and picking the lowest value from this sorted set. The resulting curve seen in Fig. 4(a) is referred to as an Initial Model Chromatogram (IMC) and the associated scaling parameter is referred to as  $s$  which operates to model a change in concentration of the standard. The  $N$  elements of the IMC are labeled by  $p_i$  where  $i = 1, \dots, N$ .



The data points seen in Fig. 4(b) are generated by taking the first derivative of the IMC. Preferably such an operation is performed with the use of a Savitzky-Golay filter as described by Abraham Savitzky and Marcel J. E. Golay in a paper entitled "Smoothing and Differentiation of Data by Simplified Least Squares Procedures"

- 5 *Analytical Chemistry*, v. 36, pp.1627-1639 (July 1964).

The  $N$  elements of the curve of Fig. 4(b) are labeled by  $p'_i$ . The scaling parameter associated with this curve is  $\delta$ . Adding the scaled elements of the curve of Fig. 4(b) to the IMC models a shift in the peak retention of the IMC by a time equal to  $t_0 = \delta/s$ .

- 10 The data points seen in Fig. 4(c) are generated by taking the element-by-element product of the curve of Fig. 4(b) with a line of unit slope and zero mean. Specifically, if there are  $N$  points within the elution time range, then  $t_i = -h, \dots, -1, 0, 1, \dots, h$  defines a curve that has unit slope and zero mean, where  $h = (N - 1)/2$ . The scaling parameter associated with this curve is  $r$ . Adding the scaled elements of the
- 15 curve of Fig. 4(c) to the IMC models a stretch in the retention times of peaks in the IMC about the IMC's midpoint. The retention time of a peak is then proportional to  $r$  times the difference between its retention time and the midpoint of the comparison range.

- The curve of Fig. 4(d) is a line having zero slope and a unit value. The scaling
- 20 parameter associated with this curve is  $b_0$ . Adding the scaled elements of the curve of Fig. 4(d) to the IMC models a change in the baseline of the IMC by an amount equal to  $b_0$ . The curve of Fig. 4(e) is a line of unit slope and zero mean. The elements of this curve are the points  $t_i$  defined above. The scaling parameter associated with this

curve is  $b_i$ . Adding the scaled elements of the curve of Fig. 4(e) to the IMC models a change in the baseline slope of the IMC by an amount equal to  $b_i$ .

Once the chromatographic variability curves are generated, the CPAS generated at 304 a design matrix with  $N$  rows and five columns. Thus in the present example, the elements of the design matrix consist of 400 rows and five columns. The elements of the design matrix are denoted by  $D_{ij}$  where  $i$  ranges from 1 to 400 and  $j$  ranges from 1 to 5. Each column of the design matrix therefore also corresponds to a scaling parameter. Each element of the design matrix is a point along one of the chromatographic variability curves. Preferably this same design matrix is associated with the standard chromatogram and is used for all values of  $k$ , which as explained above specifies an index offset.

From the design matrix ( $D$ ), a filter matrix ( $F$ ) is generated in accordance with the following relationship:

$$F = (D' D)^{-1} D' \quad (1)$$

15

The filter matrix has 5 rows and  $N$  columns. Each row of the filter matrix is an  $N$  element row-vector that acts as a projection vector associated with the respective chromatographic variability curve. In the above equation, the prime indicates matrix transposition, and the  $-1$  indicates matrix inversion. As will be understood by those skilled in the art in view of the present disclosure, the equation (1) above, advantageously implements a solution to the linear least-squares problem of fitting of the sample data to the scaled columns of the design matrix. In other words, the relationship expressed in equation (1) above, when combined with the relationships

20

shown in equations (2), (3) and (5) below provides a set of values for the five scaling parameters ( $s$ ,  $\delta$ ,  $r$ ,  $b_a$ ,  $b_f$ ) that give the lowest value of  $S_k^2$ . Alternatively, such values may be carried out by a search over a five-dimensional space, or by some iterative means, again in a five-dimensional space.

- 5           After the filter matrix is generated, a value for the offset range  $K$  is selected, and the functions shown within block 307 are performed for each index offset value  $k$ . For each index offset  $k$ , the functions within block 307 generate quantities which are indicative of the difference between the standard and sample chromatograms for that index offset  $k$ .
- 10           At step 308, best fit scaling parameters are generated as a function of the filter matrix, the sample chromatogram and the offset index  $k$ , as follows. The best fit parameters for offset index  $k$  are preferably generated by choosing absorbances from the sample chromatogram that correspond to the comparison range, offset by offset index  $k$ . Thus if the comparison range is from indices that range from  $i_{start}$  to  $i_{stop}$  then
- 15           the sample chromatogram contains elements from  $i_{start} + k$  to  $i_{stop} + k$ , which are designated by the element  $q$ . In Block 308, we solve the least-squares problem of finding the vector  $c^k$  that minimizes the quantity:

$$\|Dc^k - q\|^2 \quad (2)$$

20

The product  $Dc^k$  has the effect of weighting the columns of  $D$  with the five elements of the vector  $c^k$ . In this example,  $c^k$  is a five element column vector, whose values are the scaling factors. This product  $Dc^k$  is a model for the chromatographic data  $q^k$  based on

the chromatographic variability curves. The vector  $q^k$  contains the sample absorbances  $q_j^k$ . Finding the vector  $c^k$  that minimizes the above quantity gives a least-squares solution to the above equation provides a model that best fits the sample chromatographic data offset by  $k$ . The well known solution that gives the vector  $c^k$  that minimizes the quantity above is obtained from the filter matrix  $F$  in accordance with the following relationship:

$$c^k = F q^k \quad (3)$$

This well known formulation and solution to the least-squares problem can be found in a book by Gilbert Strang entitled "Introduction to Applied Mathematics" published by Wellesley-Cambridge Press, Wellesley, MA 02182 USA (1986) on page 37. The five elements of the column vector  $c^k$  are the least-squares estimates of each of the scaling parameters ( $s$ ,  $\delta$ ,  $r$ ,  $b_1$ ,  $b_0$ ) and are associated as follows:

$$\begin{aligned} s &= c_1^k \\ \delta &= c_2^k \\ r &= c_3^k \\ b_0 &= c_4^k \\ b_1 &= c_5^k \end{aligned} \quad (4)$$

A best-fit model for the selected index  $k$  is then generated at step 310 by generating a column vector  $m$ , with a length equal to the number of points in the comparison range (400 in the present example), in accordance with the following relationship:

$$m^k = D c^k \quad (5)$$

The effect of the matrix multiplication seen above in equation (5) is to weight the columns of  $D_g$  with the parameter values  $c^k$ , and sum the result. The resulting sum is the model  $m^k$  composing a set of 400 points which embody a model chromatogram that best fits the sample chromatogram, offset by  $k$ .

- 5           At step 312, the model chromatogram ( $m_i$ ) is compared to the sample chromatogram ( $q_i^k$ ) by obtaining the difference between the two curves to form a residual curve ( $r_i = m_i - q_i$ ). The quantity  $r_i$  is the point-to-point difference between the model and the sample. From the residual curve, at step 314, the sum of the squared residuals ( $S_i^2$ ), also termed the residual value, which measures the precision of
- 10          the fit of the sample to the model, is then generated in accordance with the following relationship:

$$S_i^2 = \sum_{i=1}^N (m_i^k - q_i^k)^2 \quad (6)$$

- 15          The right hand side of equation (6) is equivalent to the combination of equations (2) and (5).

The value  $S_i^2$  may also be normalized in accordance with the relationship shown in equation (7) below to a value  $R_i$  which expresses  $S_i^2$  as the percentage deviation between the model and the sample chromatogram:

20

$$R_i = 100 \times \sqrt{\frac{S_i^2}{\sum_{i=1}^N q_i^2}} \quad (7)$$

Once the residual value is generated at step 314, then the offset index is incremented at 314 and the steps 308-314 are repeated for each offset index in the range. The offset index value is initially chosen to be  $-k$  and is then incremented for each subsequent pass through steps 308, 310, 312 and 314 until an  $S^2$  value has been generated for each index offset value. While the offset index value is changed in Fig. 3  
5 by incrementing it, other methods of changing the value, such as decrementing from  $k$  may be used. At 318, once a value to herein  $S_k^2$  is generated for each offset index, the lowest value is selected, which is referred to herein as  $S^2$  or  $\min S^2$ . This lowest value identifies the best-fit value  $k^*$  and the parameters associated with  $k^*$ .

10 The method used to measure  $S^2$  is significant, because of the behavior of  $S^2$  when the chromatograms are of the same mixture. Because  $S^2$  measures differences after removing the effects of concentration change and chromatographic variability, a low value is obtained for  $S^2$  and one characteristic of the system residual errors.

Figs. 6(a) - 6(d) show an example of the best-fit model for an offset index of  $k$   
15 = 0. In Fig. 6(a), the IMC (solid line) and the sample chromatogram (dotted line), offset by  $k = 0$  points are shown. In Fig. 6(b) the IMC (solid line) and the best-fit model (dotted line) that best fits the sample offset by  $k = 0$  points is shown. Fig. 6(c) shows the best-fit model, (dotted line) and the sample (solid line). From these two curves a relative deviation ( $R_d$ ) is obtained of  $R_d = 15.8\%$ . Fig. 6(d) shows a plot of  
20 the residuals, which are point-to-point differences between the sample and the model chromatograms of Fig. 6(c). The sum of the squares of the values in Fig. 6(d) gives the value for  $S^2$ .

Figs. 7(a) - 7(d) are similar to Figs. 6(a) - 6(d) but show an example of the best-fit model for an offset index of  $k = 4$ . The IMC (solid line) and the sample (dotted line) are shown in Fig. 6(a), and the IMC (solid line) and the best-fit model (dotted line) are shown in Fig. 6(b). Fig. 6(c) shows the best-fit model (solid line) and the sample (dotted line). From these two curves, a relative deviation of  $R_4 = 4\%$  is obtained. As can be seen, an effect of  $k = 4$  produces a better fit than an offset of  $k = 0$ . Fig. 7(d) shows a plot of the residuals. As with Fig. 6(d), the sum of the squares of the values in this plot gives the value for  $S^2$ .

The value  $\min S^2$  is also a measure of the difference between a pair of chromatograms, whether they are from the same or different mixtures. Once such a determination of chromatographic variability has been generated from comparison of chromatograms obtained from the same or similar mixture, chromatograms of different mixtures may be compared to the standard chromatogram to determine if a different mixture is the same or different from the mixture corresponding to the standard chromatogram. Often times the standard mixture, often referred to as a reference standard or a gold standard, from which the standard chromatogram is obtained, is carefully stored, and only small amounts of the reference standard are used, perhaps on a weekly basis, as part of comparison assays. An analyst may generate one or more chromatograms from the reference standard for comparison with newly manufactured mixtures. In the simplest case, described above with reference to Figs. 1-3, two chromatograms are generated from the reference standard mixture: the standard chromatogram and the sample chromatogram. Oftentimes however, a plurality ( $N$ ) of chromatograms are generated from the reference standard as shown in Fig. 5. As seen

in Fig. 5 a plurality of standard chromatograms (standard chromatograms 1, 2, . . . ,  $M$ ) are generated from Mixture-1 which is the reference mixture. By obtaining  $M$  chromatograms from one mixture, a statistically significant distribution of  $S^2$  values can be generated and chromatograms with large deviations from the distribution values can be rejected. Other data gathering strategies are also possible. For example, in the case of clinical studies, when mixtures are from biologic sources, multiple reference mixtures may be used. In general, the problem of determining whether a value is or is not part of a distribution is a well known problem in statistics. Possible solutions to this problem are described by J.C. Miller and J.N. Miller in "Statistics for Analytical Chemistry" published by Halsted Press: a division of John Wiley & Sons, New York (1988).

As shown in Fig. 1 and explained in the accompanying description, an arbitrary number of unknown or sample mixtures may be used. In the simplest scenario, shown in Fig. 1, one reference standard mixture is used (Mixture 1), from which two chromatograms are obtained (Standard Chromatogram and Sample Chromatogram 1), and multiple unknown mixtures are used (Mixtures 2, 3, . . . ,  $N$ ) for which one chromatogram each is obtained (Sample Chromatograms 2, 3, . . . ,  $N$ ). In such a scenario, as further described in Figs. 2 and 3, a set of five variability curves, five scaling parameters, an integer offset, and a residual value are generated from the comparison of the standard chromatograms and sample chromatogram 1. The residual value  $\min S^2$  is stored, and in the following description is referred to as the reference residual. From the reference residual a threshold value which is indicative of a maximum acceptable deviation from the standard is generated by a technique



- appropriate for the application. For example, the threshold value might be some multiple of the reference residual, or alternatively, might be a fixed value. The threshold value may also be generated from an analysis of the distribution of reference residuals generated from a comparison of a plurality of sample chromatograms,
- 5 generated from the same mixture as the standard, to the standard chromatogram.

Next, the standard chromatogram is compared to one of the unknown (sample) chromatograms. Using the five variability curves generated from the standard, a new set of five scaling parameters, a new integer offset value, a new residual curve and a new residual value  $\min S^2$  is generated for each sample. Finally, the residual values

10 obtained for each of the samples (the unknown residuals) is compared to the threshold. If the unknown residual is greater than the threshold, then the mixture corresponding to that residual is determined to be different from the standard mixture. If the unknown residual is less than the threshold, then the unknown is determined to be the same (or, strictly speaking, not discernibly different from) the standard.

- 15 While the foregoing discussion has focused on an explanation of a single chromatographic comparison region obtained from a mixture, alternatives, as alluded to above exist. For example, the techniques described above could be applied to multiple, overlapping or non-overlapping comparison ranges or portions of data from each chromatogram. One mode of comparison found to be useful in the comparison of
- 20 tryptic maps is to pick a comparison range that is about five or six peak widths wide, and perform comparisons using a series of comparison ranges displaced by only a fraction of a peak width. For example, the comparison range could be 150 points, and the comparison range could be moved only by about one to five points, so each point

of data is included in approximately thirty to 150 successive comparison ranges. With such a comparison herein termed a "moving window comparison", the residual values may be plotted as a function of the center position of each comparison range. The residual values may also be combined into a single composite value. In addition, one or more scale values each derived from individual comparisons may be plotted as a function of the central point of each comparison range.

Another variable to the above description which is contemplated is the use of less than the five variability parameters reflected in the five variability curves and the five scaling parameters. In general, the advantage of retaining only a subset of the five variability parameters is that the residual difference curves and values for  $R$  and  $S^2$  are made more sensitive to those variations in the sample mixture that are reflected in the sample chromatogram. Also, the computational time needed for each comparison is reduced. This reduction can be useful in the case of moving window comparisons which are computationally intensive. However, enough parameters must be included so as to properly model the components of chromatographic variability that are present. For example, when the comparison region is short, only a few peak widths wide, it may be possible to use a model that includes only three variability parameters, with the corresponding three variability curves and the three scaling parameters. The preferred choice would be the scale factor parameterized by  $s$ , the retention time offset parameterized by  $\delta$ , and the baseline offset parameterized by  $b_0$ . Over a short comparison region, the remaining two variabilities, retention time stretch, parameterized by  $r$ , and baseline slope, parameterized by  $b_1$ , may be small enough so as not to be significant. A fourth variable consisting of either retention time stretch or

baseline slope may be added to the foregoing subset if the variations due to these effects, as produced by the chromatograph, warrant the addition.

An additional subset consists of the single variability parameter, the scale  $s$ . For this subset to be appropriate the baseline offsets and slopes must be nearly the same for each comparison region, and there must be little or no retention time stretch. Such a method may be less advantageous than the three parameter, four parameter or five parameter models described above. In particular, this method gives a value of retention time offset that equals a integral number of sample periods or index offsets. However, this method still has the advantage that it measures a retention time offset with out requiring the identification or characterization of peaks. Also, this method has the advantage of giving a value for the ratio of concentrations, described by the scale change  $s$ , a residual curve and residual value  $S^2$ . The disadvantage of the preceding one parameter method is that retention time shifts will in general not correspond to an integral number of sample periods or index offsets. We can improve the preceding method that uses the one variability parameter  $s$ , by adding to it the parameter  $\delta$  and corresponding curve associated with a retention time shift. This two-parameter method, when comparing curves that are offset by other then an integral number of sampling units, will interpolate between integral index offsets and thereby accurately measure retention time shift. In addition, this two-parameter method retains the advantage of producing a residual curve and a value for  $S^2$ . In summary, the useful subsets of the five parameters are the subset consisting of  $s$ , the subset consisting of  $s$  and  $\delta$ , the subset consisting of  $s$ ,  $\delta$ , and  $b_0$ , the subset consisting of  $s$ ,  $\delta$ ,  $b_0$  and  $r$ , and the full set of parameters,  $s$ ,  $\delta$ ,  $b_0$ ,  $r$  and  $b_I$ .

The embodiment described in connection with Figs. 1, 2 and 3 is preferably implemented as a program executing on a general purpose computer. A code listing of a preferred implementation in the MATLAB® programming language, is provided below. The below listing may be converted to executable form by interpretation via an  
5 appropriate interpreter for the MATLAB® programming language available from The Math Works, Inc., Natick, Massachusetts 01760.

Copyright 1996 Waters Corp.

```

%AlignCode Alignment demo
5 %
% [ChiSquare, BestIndOffset, SampleModel, LinearParameters, ...
% Residuals, StanOffset, Angle, PercentRSD] = ...
% PM_T_Off(StandardPattern, DerivStanPattern, StanIndToMatch, ...
% SamplePattern, IndOffset, Model)
10 %
% Inputs (all must be row vectors)
% StanIndToMatch index vector of library elements involved in match
% IndOffset vector of offset indices
%
15 %
% Outputs:
% ChiSquare Minimum for all search ranges
% BestIndOffset Best index offset corresponding to ChiSquare
% SampleModel Sample model
20 % LinearParameters Absorbance scale, Time offset, Time stretch, Baseline
Offset
% Residuals
% StanOffset Absorbance offset
% Angle Constrast angle.
25 % PercentRSD Percentage deviation between library and sample

function [ChiSquare, BestIndOffset, SampleModel, BestLinearParameters, ...
Residuals, StanOffset, Angle, PercentRSD] = ...
30 AlignCode(StandardPattern, DerivStanPattern, StanIndToMatch, ...
SamplePattern, IndOffset, Model)

% Qualification of data
[n1,d]=size(StandardPattern);
35 [n2,NumInd]=size(StanIndToMatch);
[n3,d]=size(SamplePattern);
[n4,d]=size(IndOffset);

if ~(n1==1 & n2==1 & n3==1 & n4==1)
40 error('All inputs must be row vectors')
end

% Sizes
numOffsets = length(IndOffset);
45 ChiSquareVec = zeros(1,numOffsets);
LenPattern = length(StanIndToMatch);

```

```

%Design matrix:

% Prepare Standard Pattern by selecting elements and autozeroing
5 StanPatternToMatch = StandardPattern(StanIndToMatch);
  StanOffset        = min(StanPatternToMatch);
  StanPatternToMatch = StanPatternToMatch - StanOffset;
  maxStan           = max(StanPatternToMatch);
  StanPatternToMatch = StanPatternToMatch/maxStan;
10 DerivStanPattern   = DerivStanPattern/maxStan;

% Obtain partials w/r to time and time scale
  DerivStanPatternToMatch = DerivStanPattern(StanIndToMatch);
  Ramp                    = 1:LenPattern;
15 Ramp                = Ramp - mean(Ramp);    % - LenPattern/2 to +
  LenPattern/2
  ScaleStanPatternToMatch = DerivStanPattern(StanIndToMatch) .* Ramp;

% Baseline
20 BaselineOffset = ones(size(StanPatternToMatch));

% Autozero Sample at midpoint of search range.
  SampleOffset      =
  min(SamplePattern(StanIndToMatch+floor(mean(IndOffset))));
25 SamplePattern    = SamplePattern - SampleOffset;
  SampleMax         =
  max(SamplePattern(StanIndToMatch+floor(mean(IndOffset))));
  SamplePattern      = SamplePattern/SampleMax;

30 % Design matrix
  % Five parameter model, Absorbance scale, time offset, time scale, baseline
  % offset, slope
  Design = [StanPatternToMatch', ...
            DerivStanPatternToMatch',...
            ScaleStanPatternToMatch', ...
            BaselineOffset, ...
            Ramp'];
35

40 % Least Squares solution:
  ProjMatrix = inv(Design'*Design)*Design';

% Build up SampleMatrix for each index offset
  SampleMatrix = zeros(LenPattern,numOffsets);
45 for ii = 1:numOffsets
    SampleMatrix(:,ii) = SamplePattern(StanIndToMatch+IndOffset(ii));
  end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Vectorized Computation
5 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Best fit linear parameters for each index offset.
%NumParam x numOffsets = (NumParam x LenPattern) * (LenPattern x
numOffsets)
10 LinearParameters = ProjMatrix * SampleMatrix;

% LenPattern x numOffsets = LenPat x numOff - (LenPat x Numparam)*(NumP x
numOff)
ModelCurves = Design*LinearParameters;
15 Residuals = SampleMatrix - ModelCurves;

% numOff x NumOff = (numOff x LenPat)*(LenPat x numOff);
ChiSquareVec = sum(Residuals * Residuals);

20 % Search for best Least squares solution:
[ChiSquare, iiBest] = min(ChiSquareVec);

% Report results for best fit
BestIndOffset = IndOffset(iiBest);
25 SampleModel = SampleMatrix(:,iiBest)' + SampleOffset;
BestLinearParameters = LinearParameters(:,iiBest);
Residuals = Residuals(:,iiBest);
Angle = (180/pi)*asin(sqrt(ChiSquare)/norm(SampleMatrix(:,iiBest)));
PercentRSD = 100*sqrt(ChiSquareVec/sum(StanPatternToMatch.^2));
30

```

It is to be understood that the specific mechanisms and techniques which have been described are merely illustrative of one application of the principles of the invention. Numerous modifications may be made to the methods and apparatus described without departing from the true spirit and scope of the invention.

**WHAT IS CLAIMED IS:**

1. A system for determining differences between a standard chromatogram and a sample chromatogram, said standard chromatogram represented by a set of standard data points, indicative of said standard chromatogram over a selected elution time range, said sample chromatogram represented by a set of sample data points indicative of said sample chromatogram over said elution time range, said standard data points and said sample data points being generated by sampling each respective chromatogram at a fixed rate, said system comprising:
  - 10 means for generating a plurality of sets of chromatographic variability data points from said standard data points, each of said sets of chromatographic variability data points indicative of effects of a predetermined source of chromatographic variability on said standard chromatogram; and
  - means for generating a set of modified standard data points, corresponding to  
15 said standard data points modified as a function of said chromatographic variability data points, to model chromatographic variability of a chromatograph which generates said chromatograms.
2. A system as set forth in claim 1 further comprising means, responsive to said  
20 modified standard data points and to said sample data points for generating a residual value, indicative of differences between said modified standard data points and said sample data points.



3. A system as set forth in claim 2 wherein said means for generating a set of modified standard data points comprises:

means responsive to said chromatographic variability data points and to said sample data points, for generating a plurality of scaling parameters, each of said scaling parameters corresponding to one of said sets of chromatographic variability data points; and

means for generating said modified standard data points by altering said chromatographic variability data points as a function of said scaling parameters.

10 4. A system as set forth in claim 2 wherein said means for generating a set of modified standard data points comprises:

means, responsive to a selected offset index range, indicative of an amount by which said standard data points may be shifted, for generating a plurality of sets of said scaling parameters as a function of said chromatographic variability data points and a corresponding set of standard data points obtained from said standard data points shifted by an offset index value within said offset index range;

means for generating a plurality of sets of said modified standard data points by altering said chromatographic variability data points as a function of each of said sets of said scaling parameters;

20 means for generating said residual value for each of said sets of said modified standard data points;

means for selecting said set of modified sample data points as a function of the set of modified standard data points having the lowest residual value.

5. A system as set forth in claim 4 wherein said scaling parameters comprise:
- a first scaling parameter indicative of a change in concentration attributable to variability of a chromatograph used to generate said standard and sample
- 5 chromatograms;
- a second scaling parameter indicative of a shift in peak retention attributable to variability of said chromatograph used to generate said standard and sample chromatograms;
- a third scaling parameter indicative of a stretch in peak retention time
- 10 attributable to variability of said chromatograph used to generate said standard and sample chromatograms;
- a fourth scaling parameter indicative of a change in baseline attributable to variability of said chromatograph used to generate said standard and sample chromatograms; and
- 15 a fifth scaling parameter indicative of a change in baseline slope attributable to variability of said chromatograph used to generate said standard and sample chromatograms.
6. A system as set forth in claim 4 wherein said scaling parameters comprise:
- 20 a first scaling parameter indicative of a change in concentration attributable to variability of a chromatograph used to generate said standard and sample chromatograms.

7. A system as set forth in claim 6 wherein said scaling parameters comprise further comprise:

a second scaling parameter indicative of a shift in peak retention attributable to variability of said chromatograph used to generate said standard and sample

5 chromatograms.

8. A system as set forth in claim 7 wherein said scaling parameters comprise further comprise:

a third scaling parameter indicative of a change in baseline attributable to  
10 variability of said chromatograph used to generate said standard and sample chromatograms.

9. A system as set forth in claim 8 wherein said scaling parameters comprise further comprise:

15 a fourth scaling parameter indicative of a stretch in peak retention time attributable to variability of said chromatograph used to generate said standard and sample chromatograms.

10. A system as set forth in claim 1 wherein said means for generating a plurality of  
20 sets of chromatographic variability data points from said standard data points comprises:

means for generating a first set of chromatographic variability data points by subtracting a minimum absorbance in said standard chromatogram from said points of

said standard chromatogram, said first set of chromatographic variability data points indicative of a change in concentration attributable to variability of said chromatograph used to generate said standard and sample chromatograms;

means for generating a second set of chromatographic variability data points by  
5 obtaining a first derivative of said first set of chromatographic variability data points, said second set of chromatographic variability data points indicative of a change in a shift of peak retention of said first set of chromatographic variability data points;

means for generating a third set of chromatographic variability data points by multiplying said second set of chromatographic variability data points with a line  
10 characterized by a unit slope and a zero mean, said third set of chromatographic variability data points indicative of a stretch, about a midpoint, in the retention time of peaks in said first set of chromatographic variability data points;

means for generating a fourth set of chromatographic variability data points by generating a line which has a unit value, said fourth set of chromatographic variability  
15 data points indicative of an offset in a baseline of said first set of chromatographic variability data points; and

means for generating a fifth set of chromatographic variability data points by generating a line characterized by a unit slope and a zero mean, said fifth set of chromatographic variability data points indicative of a change in baseline slope of said  
20 first set of chromatographic variability data points.

11. A system as set forth in claim 5 wherein said means for generating a plurality of sets of chromatographic variability data points from said standard data points comprises:

means for generating a first set of chromatographic variability data points by

- 5 subtracting a minimum absorbance in said standard chromatogram from said points of said standard chromatogram, said first set of chromatographic variability data points indicative of a change in concentration attributable to variability of said chromatograph used to generate said standard and sample chromatograms;

means for generating a second set of chromatographic variability data points by

- 10 obtaining a first derivative of said first set of chromatographic variability data points, said second set of chromatographic variability data points indicative of a change in a shift of peak retention of said first set of chromatographic variability data points;

means for generating a third set of chromatographic variability data points by

- 15 multiplying said second set of chromatographic variability data points with a line characterized by a unit slope and a zero mean, said third set of chromatographic variability data points indicative of a stretch, about a midpoint, in the retention time of peaks in said first set of chromatographic variability data points;

means for generating a fourth set of chromatographic variability data points by

- generating a line which has a unit value, said fourth set of chromatographic variability  
20 data points indicative of an offset in a baseline of said first set of chromatographic variability data points; and

means for generating a fifth set of chromatographic variability data points by

generating a line characterized by a unit slope and a zero mean, said fifth set of

chromatographic variability data points indicative of a change in baseline slope of said first set of chromatographic variability data points.

12. A system as set forth in claim 4 wherein said standard chromatogram and said sample chromatogram are each generated from the same mixture.

13. A system as set forth in claim 4 wherein said standard chromatogram and said sample chromatogram are each generated from different mixtures.

14. A system for determining chromatographic variability between a standard chromatogram and a sample chromatogram each of which is represented by a set of data points indicative of said respective chromatogram over a selected elution time range, each set of said data points being generated by sampling each respective chromatogram at a fixed rate, said system comprising:

means for generating, as a function of said data points corresponding to said standard chromatogram and as a function of said data points corresponding to said sample chromatogram, a plurality of scaling parameters as a function of changes in said standard chromatogram over said elution time range, said scaling parameters indicative of chromatographic variability;

means, responsive to said scaling parameters and to said data points corresponding to said standard chromatogram, for generating a best-fit model of said sample chromatogram which reflects said sample chromatogram modified to remove chromatographic variability; and

means, responsive to said best-fit model, for generating a residual difference value indicative of a difference between said standard chromatogram and said sample chromatogram as modified to remove said chromatographic variability.

15. A method of determining chromatographic variability comprising:
- obtaining from two chromatograms, which are each obtained from the same mixture, a set of data points from each chromatogram, said data points obtained by sampling said respective chromatogram at equal intervals over a comparison range including at least two peaks; and
- estimating, without characterization of chromatographic features associated with a peak, at least two effects of chromatographic variability manifested in said comparison range.
16. A method as set forth in claim 15 comprising the further step of estimating, without identification of chromatographic features associated with a peak, a change in retention time scale between the comparison ranges of said two chromatograms.
17. A method as set forth in claim 15 comprising the further step of generating a comparison factor indicative of dissimilarity between the comparison ranges of said two chromatograms.
18. A method as set forth in claim 16 wherein said retention time estimation and said retention time scale change estimation are each performed by a linear least-squares

procedure which minimizes a sum of squared residual differences between the comparison ranges of said two chromatograms.

19. A method as set forth in claim 16 comprising the further step of estimating,  
5 without identification of chromatographic features associated with a peak, a change in concentration between the comparison ranges of said two chromatograms.
20. A method as set forth in claim 16 comprising the further step of estimating,  
without identification of chromatographic features associated with a peak, a change in  
10 retention time offset between the comparison ranges of said two chromatograms.
21. A method as set forth in claim 16 comprising the further step of estimating,  
without identification of chromatographic features associated with a peak, a change in  
baseline slope between the comparison ranges of said two chromatograms.  
15
22. A method as set forth in claim 16 comprising the further step of estimating,  
without identification of chromatographic features associated with a peak, a change in  
retention time stretch between the comparison ranges of said two chromatograms.
- 20 23. A method as set forth in claim 16 comprising the further step of estimating,  
without identification of chromatographic features associated with a peak, a change in  
retention time offset between the comparison ranges of said two chromatograms.



24. A system as set forth in claim 3 further comprising means for dividing said chromatographic variability data points and said sample data points into a plurality of comparison ranges and estimating, without characterization of chromatographic features associated with a peak, at least two effects of chromatographic variability manifested in each comparison range.
25. A system as set forth in claim 24 wherein said comparison ranges are overlapping.
26. A system as set forth in claim 24 wherein said comparison ranges are non-overlapping.
27. A system as set forth in claim 25 wherein said comparison range is displaced by one or more sample points.
28. The system of claim 24 wherein one or more scale values and residual values from a comparison region are plotted as a function of the center of each comparison region.
29. A method for comparing a standard chromatogram and a sample chromatogram, said standard chromatogram represented by a set of standard data points, indicative of said standard chromatogram over a selected elution time range, said sample chromatogram represented by a set of sample data points indicative of said sample chromatogram over said elution time range, said standard data points and said sample data points being generated by sampling each respective chromatogram at a fixed rate, said method comprising:
- generating a plurality of sets of chromatographic variability data points from said standard data points, each of said sets of chromatographic variability data points indicative of effects of a predetermined source of chromatographic variability on said standard chromatogram;

- dividing said chromatographic variability data points and said sample data points into a plurality of comparison ranges and estimating, without characterization of chromatographic features associated with a peak, at least two effects of chromatographic variability manifested in each range;
- 5 generating a plurality of scaling parameters, each of said scaling parameters corresponding to one of said sets of chromatographic variability data points;
- generating a set of modified standard data points by altering said chromatographic variability data points as a function of said scaling parameters;
- 10 generating a best fit model for said sample data points with respect to said modified standard data points in one or more comparison ranges; and,
- generating a residual value, indicative of differences between said modified standard data points and said sample data points.
- 15 30. The method of claim 29 wherein a second sample is used to generate a second chromatogram and the method comprises:
- generating a best fit model with respect to said modified standard chromatogram; and,
- generating a residual value, indicative of differences between said modified
- 20 standard data points and said second sample data points.

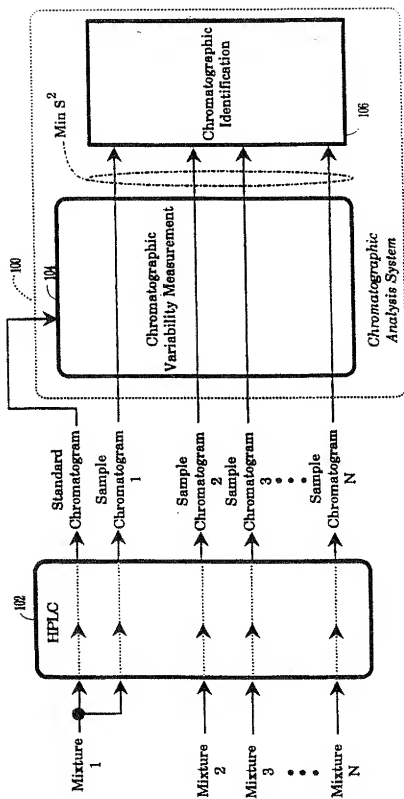
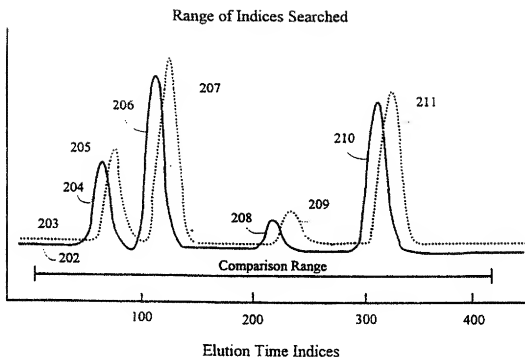


Fig. 1

**Fig. 2**

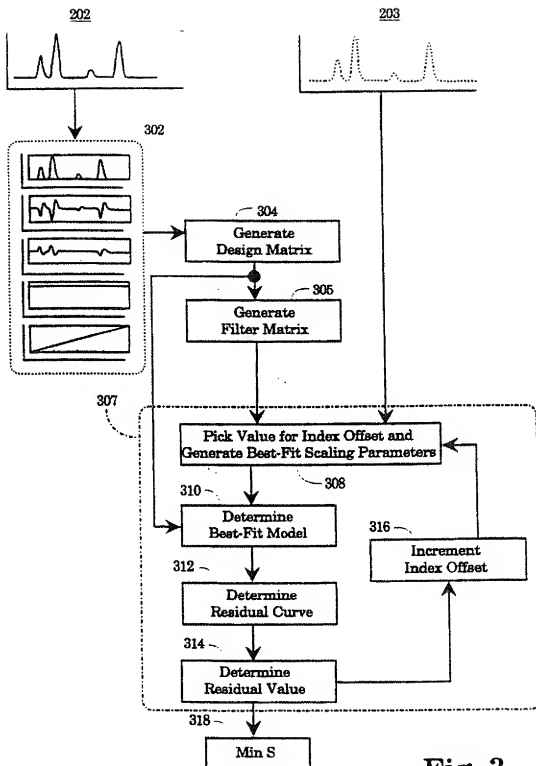
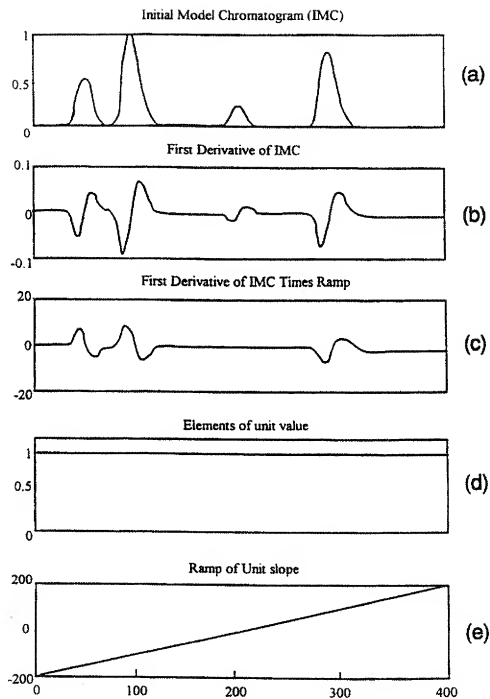


Fig. 3

**Fig. 4(a-e)**

4/7

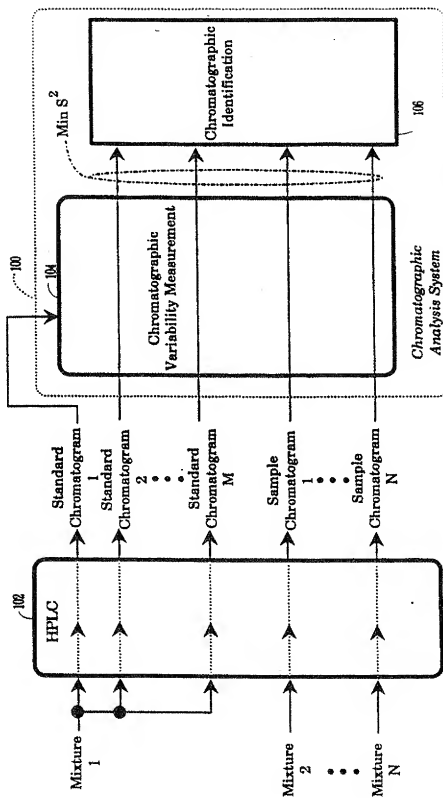
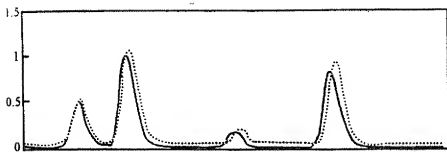
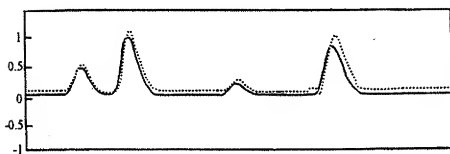
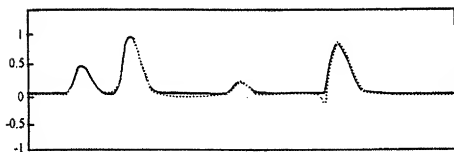
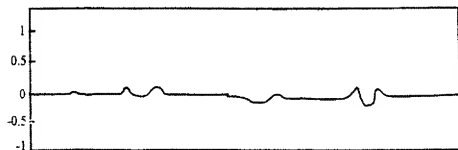
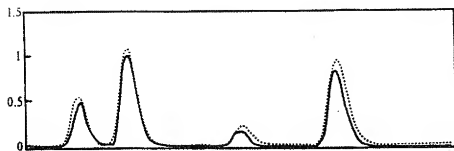
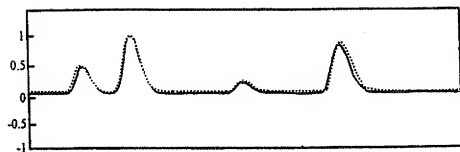
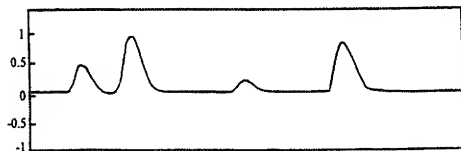
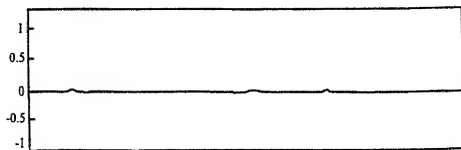


Fig. 5

**Fig. 6(a)****Fig. 6(b)****Fig. 6(c)****Fig. 6(d)**



**Fig. 7(a)****Fig. 7(b)****Fig. 7(c)****Fig. 7(d)**

# INTERNATIONAL SEARCH REPORT

In- vent Application No  
PCT/US 97/06135

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 G01N30/86 G06F15/20

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G01N G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 665 433 A (WYATT TECHNOLOGY) 2 August 1995 see page 8, line 10-47 ---	1-4, 14
A	CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, vol. 22, no. 1, 1 January 1994, pages 49-61, XP000414576 ANDERSSON R ET AL: "SIMPLEX FOCUSING OF RETENTION TIMES AND LATENT VARIABLE PROJECTIONS OF CHROMATOGRAPHIC PROFILES" see the whole document ---	1-30
A	WO 93 21592 A (DOW CHEMICAL CO) 28 October 1993 see the whole document ---	1-30
-/--		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \* "A" document defining the general state of the art which is not considered to be of particular relevance
- \* "E" earlier document but published on or after the international filing date
- \* "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \* "O" document referring to an oral disclosure, use, exhibition or other means
- \* "P" document published prior to the international filing date but later than the priority date claimed

- \* "I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \* "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \* "&" document member of the same patent family

Date of the actual completion of the international search  16 July 1997	Date of mailing of the international search report  - 1. 08. 97
Name and mailing address of the ISA European Patent Office, P.O. 5818 Patentamt NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer  Mueller, T

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 97/06135

## C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4 740 903 A (NAKATSUKA KIYOHARU ET AL) 26 April 1988 see the whole document -----	1-30

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 97/06135

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0665433 A	02-08-95	US 5528365 A JP 8035961 A	18-06-96 06-02-96
WO 9321592 A	28-10-93	CA 2133412 A EP 0636261 A JP 8500898 T US 5592402 A	28-10-93 01-02-95 30-01-96 07-01-97
US 4740903 A	26-04-88	JP 7018853 B JP 60239669 A DE 3584307 A EP 0161672 A	06-03-95 28-11-85 14-11-91 21-11-85